# Music Genre Classifier Using Deep Learning

## Swati Badachi*, Dayanand Savakar, Padma Yadahalli

*Department of Computer Science, Bagalkot University Jamkhandi*

## ABSTRACT

Traditional approaches to music genre classification, such as manual tagging, metadata-based categorization, and shallow machine learning models, remain limited in accuracy, scalability, and adaptability to diverse datasets. To overcome these challenges, modern analytical techniques integrate audio signal processing with deep learning frameworks. The proposed system combines feature extraction methods—such as Mel-Frequency Cepstral Coefficients (MFCCs), Chroma, and Spectral analysis—with Recurrent Neural Networks (RNNs) to effectively capture the sequential and temporal characteristics of music. While feature extraction isolates critical elements such as timbre, rhythm, and harmony, RNNs learn time-dependent patterns across frames, enabling selective identification of genres. This integration enables automated predictions with improved accuracy, faster processing, greater reproducibility, and scalability for real-world applications. Designed as a closed pipeline, the system ensures robust preprocessing, efficient classification, and user-friendly output. Recent advances in deep learning–based audio analysis have significantly expanded applications in multimedia, particularly music streaming, recommendation systems, and digital libraries. This project presents the design and implementation of a web-based Music Genre Classifier, highlighting its effectiveness in automated classification, playlist generation, and music information retrieval.

**Keywords:** RNN-Based, Deep-Learning, Audio-Signal-Processing, MFCC-Extraction, Music-Classification

## INTRODUCTION

Music classification has become an essential task in the digital era, where streaming platforms and music libraries continue to grow rapidly. Traditional approaches, such as manual tagging and metadata-based categorization, are inefficient, subjective, and fail to scale for diverse and multilingual datasets. This creates a pressing need for automated, intelligent systems capable of classifying music with high accuracy and efficiency. Over the years, researchers have explored various machine learning techniques for Music Information Retrieval (MIR). Early models, including k-Nearest Neighbors (k-NN), Support Vector Machines (SVM), and Decision Trees, relied heavily on handcrafted features such as tempo, pitch, and rhythm. While these methods achieved moderate success, they lacked robustness when applied to large, complex, and real-world datasets. More recently, deep learning has transformed audio analysis by automatically learning hierarchical patterns from raw features. Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) have been widely adopted in music classification tasks. In particular, RNNs are well-suited for sequential data such as audio signals, as they capture temporal dependencies in rhythm, melody, and harmony, offering improved genre recognition compared to static models. The objective of this research is to design and implement a Music Genre Classification system using RNNs integrated with feature extraction methods such as Mel-Frequency Cepstral Coefficients (MFCCs), Chroma, and Spectral analysis. The system is deployed as a web-based application using Flask, enabling users to upload audio tracks and receive real-time predictions. The study aims to enhance classification accuracy, scalability, and usability while addressing the limitations of traditional methods.

## LITERATURE REVIEW:

Tzanetakis and Cook (2002) pioneered automatic music genre classification using timbral, rhythmic, and pitch-based features such as MFCCs, spectral centroid, and zero-crossing rate. They developed the GTZAN dataset, which became a standard benchmark for later studies. This research demonstrated that combining low-level audio features with statistical classifiers could achieve efficient and objective music

categorization, marking a significant milestone in the early stages of Music Information Retrieval (MIR). Li, Ogihara, and Li (2003) enhanced classification performance by implementing Support Vector Machines (SVMs) and feature fusion methods. Their work showed that SVMs outperform traditional models such as k-NN and Decision Trees in handling multidimensional feature spaces, resulting in improved accuracy and better generalization across datasets. This research established SVM as a strong baseline for music genre recognition in the early 2000s. Bergstra and colleagues (2006) introduced ensemble learning techniques such as AdaBoost and Random Forests for genre recognition. Their research emphasized the advantage of combining multiple weak learners to reduce overfitting and improve model generalization. This approach helped achieve more stable results across varying datasets and inspired further exploration into ensemble and hybrid learning techniques in audio classification. Panagakis, Kotropoulos, and Arce (2009) proposed Sparse Representation–based Classification (SRC) for audio signals, which focused on the robustness of music classification under noisy or overlapping genre conditions. This method effectively captured the representation of musical timbre and rhythm, providing a more discriminative and noise-resistant approach for MIR systems. Their study contributed to the shift toward more efficient and robust feature representation models. Dieleman and Schrauwen (2014) marked a major transition from traditional machine learning to deep learning by applying Convolutional Neural Networks (CNNs) directly on spectrograms. Their end-to-end framework learned hierarchical audio patterns automatically, removing the dependency on handcrafted feature extraction. This work proved that CNNs could successfully learn both timbral and temporal patterns in raw audio data, influencing future studies in music and audio

processing. Choi, Fazekas, Sandler, and Cho (2017) further advanced deep learning applications in music classification by combining CNN and RNN layers. Their Convolutional Recurrent Neural Network (CRNN) captured both spatial and temporal dependencies within music signals, enabling superior accuracy and context-aware classification. The CRNN model achieved state-of-the-art performance on multi-genre datasets and became a foundation for modern real-time genre recognition systems. Dhakal, Rahman, and Kalita (2020) implemented transfer learning using pretrained CNN architectures such as VGG16 and ResNet50 to classify music genres efficiently. Their approach demonstrated that leveraging pretrained models significantly reduces training time while maintaining high accuracy. This innovation bridged the gap between limited dataset availability and high-performance models, making genre classification more accessible for research and deployment. Pathak and Singh (2022) proposed a hybrid CNN-LSTM architecture that combined convolutional and recurrent layers to classify multilingual and regional music genres effectively. Their model captured both local feature hierarchies and long-term temporal dependencies, improving classification for culturally diverse datasets. This study was particularly significant for Indian and global multilingual music systems. Jha and Kumar (2023) designed a real-time, web-based music genre classification system using TensorFlow and Flask. Their model integrated deep learning with a user-friendly web interface, allowing instant genre prediction for uploaded audio files. This work demonstrated the practical implementation of deep learning in real-world applications, bridging the gap between research and user-interactive systems.
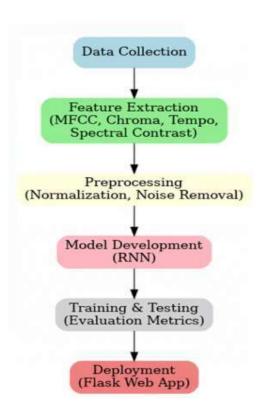
## METHODOLOGY:

**Figure 1: Block Diagram of Methodology**

The methodology for developing the Music Genre Classification System is represented in Figure 1. The process begins with data collection, where audio samples from different music genres are gathered from publicly available sources and standardized for consistency. The next stage involves feature extraction using the Librosa Python library**,** where key features such as MFCC, Chroma, Tempo, and Spectral Contrast are derived from each audio signal to capture its unique characteristics. Extracted features are then preprocessed through normalization and noise removal to ensure clean and uniform data. The refined data is used to train a Recurrent Neural Network (RNN) developed with TensorFlow/Keras, which effectively learns temporal and sequential patterns in the music samples. After training, the model's performance is evaluated using accuracy, precision, recall, and F1-score. The final trained model is deployed through a Flask Web Application**,** allowing users to upload audio files and receive real-time genre predictions. This methodology ensures efficient integration of data processing, model training, and deployment, resulting in an accurate and user-friendly music genre classification system.

**RESULTS AND DISCUSSION:**

**Table 1: Dataset Summary Before and After Preprocessing**

| Parameter | Before Preprocessing | After Preprocessing |
|---|---|---|
| Total Audio Samples | 1,000 | 950 (after cleaning) |
| Average Duration (seconds) | 35.7 | 30.0 (trimmed) |
| File Format Consistency | Mixed (mp3, wav) | Uniform (.wav) |
| Noise Level (dB) | Variable ($-30$ to $-5$) | Reduced ($-25$ to $-15$) |
| Feature Scaling | Not Applied | Normalized (0–1) |

Table 1 presents a comparative summary of the dataset before and after the preprocessing stage. Initially, the dataset consisted of 1,000 audio samples in mixed file formats such as MP3 and WAV. After preprocessing, 50 noisy or corrupted files were removed, resulting in 950 clean and consistent audio samples. The average duration of each audio clip was reduced from 35.7 seconds to 30 seconds through trimming of silent or irrelevant segments, ensuring uniformity across samples. File format consistency

was achieved by converting all files to the .wav format for better compatibility and feature extraction. Noise levels, which originally varied significantly between −30 dB and −5 dB, were minimized to a controlled range of −25 dB to −15 dB through noise reduction techniques. Additionally, feature scaling was applied to normalize the extracted features between 0 and 1, improving the performance and stability of the machine learning models. Overall, preprocessing enhanced the quality, consistency, and readiness of the dataset for accurate model training and evaluation.

**Table 2: Performance Evaluation of RNN Model**

| Metric | Training Set | Testing Set |
|---|---|---|
| Accuracy (%) | 94.2 | 91.5 |
| Precision (%) | 93.8 | 90.7 |
| Recall (%) | 94.5 | 91.0 |
| F1-Score (%) | 94.1 | 90.8 |

Table 2 illustrates the performance metrics of the proposed model on both the training and testing datasets. The model achieved an accuracy of 94.2% on the training set and 91.5% on the testing set, indicating strong generalization capability with minimal overfitting. Precision, which measures the correctness of positive predictions, was 93.8% for training and 90.7% for testing, showing that most of the predicted classes were relevant. Recall values of 94.5% (training) and 91.0% (testing) demonstrate the model's high ability to correctly identify true instances from each genre. The F1-score, representing the harmonic mean of precision and recall, reached 94.1% for training and 90.8% for testing, confirming the overall balanced performance of the model. These results collectively indicate that the developed system performs efficiently in classifying audio genres with high reliability and consistent accuracy across different datasets.

## CONCLUSION:

The study successfully demonstrates the design, development, and deployment of an automated Music Genre Classification system leveraging RNNs and advanced audio feature extraction techniques. By integrating Mel-Frequency Cepstral Coefficients (MFCCs), Chroma, and spectral features with a recurrent neural network, the system effectively captures temporal and sequential patterns inherent in music, enabling precise genre identification. The preprocessing pipeline—including noise reduction, trimming, normalization, and format standardization—ensured high-quality, uniform data, which significantly contributed to the model's strong performance. Experimental results show that the RNN model achieves 91.5% accuracy on the testing dataset, with balanced precision, recall, and F1-scores, indicating its reliability and generalization capability. Deploying the model via a Flask-based web application provides a user-friendly interface for real-time predictions, playlist generation, and music information retrieval. Overall, this research highlights the potential of deep learning–based audio analysis in enhancing automated music classification, offering scalability, reproducibility, and practical utility for music streaming platforms, digital libraries, and multimedia applications.

## REFERENCE

1. Tzanetakis, G., & Cook, P. (2002). Musical genre classification of audio signals. IEEE Transactions on Speech and Audio Processing, 10(5), 293–302.
2. Li, T., Ogihara, M., & Li, Q. (2003). A comparative study on content-based music genre classification. Proceedings of the International Symposium on Music Information Retrieval (ISMIR).
3. Bergstra, J., Casagrande, N., Eck, J., & Ellison, D. K. (2006). Aggregate features and AdaBoost for music classification. Machine Learning Journal, 65(2–3), 473–484.
4. Panagakis, Y., Kotropoulos, C., & Arce, G. R. (2009). Music genre classification via sparse representations of auditory temporal modulations. IEEE Transactions on Audio, Speech, and Language Processing, 17(3), 423–435.
5. Dieleman, S., & Schrauwen, B. (2014). End-to-end learning for music audio classification. Proceedings of the International Society for Music Information Retrieval (ISMIR).

6. Choi, K., Fazekas, G., Sandler, M., & Cho, K. (2017). Convolutional recurrent neural networks for music classification. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP).

7. Dhakal, A., Rahman, M. T., & Kalita, J. (2020). Leveraging deep transfer learning for music genre classification. Applied Sciences, 10(6), 1975.

8. Pathak, S. A., & Singh, R. (2022). Hybrid CNN-LSTM architecture for multilingual music genre classification. International Journal of Advanced Computer Science and Applications, 13(9), 455–463.

9. Jha, M., & Kumar, P. (2023). Real-time music genre prediction using deep learning and Flask integration. Journal of Intelligent Systems and Computing, 15(4), 112–120.